Guidance for Industry
Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics

DRAFT GUIDANCE

This guidance document is being distributed for comment purposes only.

Comments and suggestions regarding this draft document should be submitted 60 days of publication in the *Federal Register* of the notice announcing the availability of the draft guidance. Submit comments to the Division of Dockets Management (HFA-305), Food and Drug Administration, 5630 Fishers Lane, rm. 1061, Rockville, MD 20852. All comments should be identified with the docket number listed in the notice of availability that publishes in the *Federal Register*.

For questions regarding this draft document contact (CDER drugs) Grant Williams at 301-594-5758, (CDER biologics) Patricia Keegan at 301-827-5097, or (CBER biologics) Steven Hirschfeld at 301-827-6536.

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)

April 2005
Clinical/Medical

Guidance for Industry
Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics

Additional copies are available from:

Office of Training and Communications
Division of Drug Information, HFD-240
Center for Drug Evaluation and Research
Food and Drug Administration
5600 Fishers Lane
Rockville, MD 20857
(Tel) 301-827-4573
http://www.fda.gov/cder/guidance/index.htm

or

Office of Communication, Training, and
Manufacturers Assistance, HFM-40
Center for Biologics Evaluation and Research
Food and Drug Administration
1401 Rockville Pike, Rockville, MD 20852-1448
(Tel) 800-835-4709 or 301-827-1800
http://www.fda.gov/cber/guidelines.htm

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)

April 2005
Clinical/Medical

TABLE OF CONTENTS

I. INTRODUCTION 1
II. BACKGROUND 2
  A. Regulatory Requirements for Effectiveness 2
  B. Endpoints Supporting Past Approvals in Oncology 3
III. GENERAL ENDPOINT CONSIDERATIONS 4
  A. Overall Survival 5
  B. Endpoints Based on Tumor Assessments 6
    1. *Disease-Free Survival* *7*
    2. *Objective Response Rate* *8*
    3. *Time to Progression and Progression-Free Survival* *8*
      a. TTP vs. PFS 9
      b. PFS as an endpoint to support drug approval 9
      c. PFS trial design issues 10
      d. Analysis of PFS 10
      e. Future methods for assessing progression 11
    4. *Time to Treatment Failure* *11*
  C. Endpoints Involving Symptom Assessment 12
    1. *Specific Symptom Endpoints* *12*
    2. *Problems Encountered with Symptom Data* *13*
  D. Biomarkers 13
IV. ENDPOINTS AND CLINICAL TRIAL DESIGN; SELECTED ISSUES 14
  A. Single-Arm Studies 14
  B. Studies Designed to Demonstrate Noninferiority 15
  C. No Treatment or Placebo Control 16
  D. Isolating Drug Effect in Combinations 16
  E. Trial Designs for Radiotherapy Protectants and Chemotherapy Protectants 16
V. SUMMARY AND CONCLUSION 16
APPENDIX 1: THE COLLECTION OF TUMOR MEASUREMENT DATA 18
APPENDIX 2: ISSUES TO CONSIDER IN PFS ANALYSIS 19
APPENDIX 3: EXAMPLE TABLES FOR PFS ANALYSIS 21
APPENDIX 4: INDEPENDENT REVIEW OF TUMOR ENDPOINTS 23

Guidance for Industry[1]
Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics

This draft guidance, when finalized, will represent the Food and Drug Administration's (FDA's) current thinking on this topic. It does not create or confer any rights for or on any person and does not operate to bind FDA or the public. You can use an alternative approach if the approach satisfies the requirements of the applicable statutes and regulations. If you want to discuss an alternative approach, contact the FDA staff responsible for implementing this guidance. If you cannot identify the appropriate FDA staff, call the appropriate number listed on the title page of this guidance.

I. INTRODUCTION

This guidance provides recommendations to sponsors on endpoints for cancer clinical trials submitted to the FDA to support effectiveness claims in new drug applications (NDAs), biologics license applications (BLAs), or supplemental applications.[2]

The FDA is developing guidance on oncology endpoints through a process that includes public workshops of oncology experts and discussions before the FDA's Oncologic Drugs Advisory Committee (ODAC).[3] This guidance is the first in a planned series of cancer endpoint guidances. It provides background information and discusses general regulatory principles. Each subsequent guidance document will focus on endpoints for specific cancer types (e.g., lung cancer, colon cancer) to support drug approval or labeling claims. The endpoints discussed in this guidance document are for drugs to treat patients with an existing cancer. This guidance does not address endpoints for drugs to prevent or decrease the incidence of cancer.

FDA's guidance documents, including this guidance, do not establish legally enforceable responsibilities. Instead, guidances describe the Agency's current thinking on a topic and should be viewed only as recommendations, unless specific regulatory or statutory requirements are cited. The use of the word *should* in Agency guidances means that something is suggested or recommended, but not required.

---

[1] This guidance has been prepared by the Division of Oncology Drug Products and the Division of Therapeutic Biologic Oncology Drug Products in the Center for Drug Evaluation and Research (CDER) in cooperation with the Center for Biologics Evaluation and Research (CBER) at the Food and Drug Administration.

[2] For the purposes of this guidance, all references to *drugs* include both human drugs and biological products unless otherwise specified.

[3] Transcripts are available at http://www.fda.gov/cder/drug/cancer_endpoints/default.htm.

II. BACKGROUND

Clinical trial endpoints serve different purposes. In conventional oncology drug development, early phase clinical trials evaluate safety and identify evidence of biological drug activity, such as tumor shrinkage. Endpoints for later phase efficacy studies evaluate whether a drug provides a clinical benefit such as prolongation of survival or an improvement in symptoms. The following sections discuss the general regulatory requirements for efficacy and how they have influenced endpoint selection for the approval of cancer drugs. Later sections describe these endpoints in more detail and discuss whether they might serve as measures of disease activity or clinical benefit in various clinical settings.

A. Regulatory Requirements for Effectiveness

The requirement that new drugs show effectiveness is based on a 1962 amendment to the Federal Food, Drug, and Cosmetic Act. This law requires substantial evidence of effectiveness and specifies that this evidence must be derived from adequate and well-controlled clinical investigations. Clinical benefits that have supported drug approval have included important clinical outcomes (e.g., increased survival, symptomatic improvement) but have also included effects on established surrogate endpoints (e.g., blood pressure or serum cholesterol).

In 1992, the accelerated approval regulations (21 CFR part 314, subpart H and 21 CFR part 601, subpart E) allowed use of additional endpoints for approval of drugs or biological products that are intended to treat serious or life-threatening diseases and that either demonstrate an improvement over available therapy or provide therapy where none exists. In this setting, the FDA may grant approval based on an effect on a surrogate endpoint that is *reasonably likely* to predict clinical benefit (“based on epidemiologic, therapeutic, pathophysiologic, or other evidence”). These surrogates are less well-established than surrogates in regular use, such as blood pressure or cholesterol for cardiovascular disease. A drug is approved under the accelerated approval regulations on condition that the manufacturer conduct clinical studies to verify and describe the actual clinical benefit. If the postmarketing studies fail to demonstrate clinical benefit or if the applicant does not demonstrate due diligence in conducting the required studies, the drug may be removed from the market under an expedited process. From December 1992 to June 2004, 22 cancer drug applications were approved under the accelerated approval regulations. In the following discussion, we will use the term *regular approval* to designate the longstanding route of drug approval based on demonstrating clinical benefit to distinguish it from *accelerated approval* associated with use of a surrogate endpoint that is reasonably likely to predict benefit.

The nature of evidence to support drug approval, including the preferred number of clinical trials, is discussed in general FDA guidance documents. In most cases, the FDA has recommended at least two well-controlled clinical trials. In some cases, the FDA has found that evidence from a single trial was sufficient, but generally only in cases in which a single multicenter study provided highly reliable and statistically strong evidence of an important clinical benefit, such as an effect on survival, and in which confirmation of the result in a second trial would be practically or ethically impossible.[4] For drugs approved for treatment of patients with a specific stage of a particular malignancy, evidence from one trial may be sufficient to support an efficacy supplement for treatment of a different stage of the same cancer.[5]

B. Endpoints Supporting Past Approvals in Oncology

For regular approval, it is critical that the sponsor show direct evidence of clinical benefit or improvement in an established surrogate for clinical benefit. In oncology, survival is the gold standard for clinical benefit, but the FDA has accepted other endpoints for cancer drug approval. Indeed, in the 1970s the FDA usually approved cancer drugs based on objective response rate (ORR), determined by tumor assessments from radiologic tests or physical exam. In the early 1980s, after discussion with the ODAC, the FDA determined that it would be more appropriate for cancer drug approval to be based on more direct evidence of clinical benefit, such as improvement in survival or in a patient's quality of life (QOL), improved physical functioning, or improved tumor-related symptoms — benefits not always predicted by ORR.

Over the next decade, several endpoints were used as surrogates for benefit. Improvement in disease-free survival supported drug approval in selected surgical adjuvant settings (when a large proportion of patients had cancer symptoms at the time of recurrence). Durable complete response was considered an acceptable endpoint in testicular cancer and acute leukemia (a de facto improvement in survival because the untreated conditions were quickly lethal) and in some chronic leukemias and lymphomas (where it was clear that remission would lead to less infection, bleeding, and blood product support). The FDA has also considered that a very high ORR alone might sometimes support regular approval, but that response duration, relief of tumor-related symptoms, and drug toxicity should also be considered (O'Shaughnessy and Wittes et al., 1991, Commentary Concerning Demonstration of Safety and Efficacy of Investigational Anticancer Agents in Clinical Trials, J Clin Oncol 9:2225-2232). ORR has been an especially important endpoint for the less toxic drugs, such as the hormonal drugs for breast cancer, where improvement in this endpoint has been the basis for regular approval. Improvement in tumor-related symptoms in conjunction with an improved ORR and an adequate response duration supported approval in several clinical settings.

In the last decade, in addition to its limited role in regular approval, ORR has been the primary surrogate endpoint used to support cancer drug accelerated approval for several reasons. First, ORR is directly attributable to drug effect (tumors rarely shrink spontaneously and, therefore, ORR can be accurately assessed in single-arm studies). Second, tumor response is widely accepted as relevant by oncologists and has a long-accepted role in guiding cancer treatment. Finally, if the ORR is high enough and the responses are of sufficient duration, ORR does indeed seem *reasonably likely* to predict clinical benefit.

---

[4] See guidance for industry *Providing Clinical Evidence of Effectiveness for Human Drug and Biological Products* (http://www.fda.gov/cder/guidance/index.htm)

[5] See guidance for industry *FDA Approval of New Cancer Treatment Uses for Marketed Drug and Biological Products* (http://www.fda.gov/cder/guidance/index.htm)

Drugs approved under accelerated approval regulations must provide a benefit over available therapy. To satisfy this requirement, many sponsors have designed single-arm studies in patients with refractory tumors where, by definition, no available therapy exists.

III. GENERAL ENDPOINT CONSIDERATIONS

The following is an overview of general issues in cancer drug development. A discussion of commonly used cancer endpoints is followed by a discussion of pertinent issues in cancer clinical trial design using these endpoints. Future guidance documents will discuss these issues in more detail with regard to specific treatment indications. Endpoints that will be discussed include overall survival, endpoints based on tumor assessments (e.g., disease-free survival, ORR, time to progression, progression-free survival, time to treatment failure), and endpoints based on symptom assessment. A comparison of important endpoints in cancer drug approval is provided in Table 1. Many of the issues relating to the proper analysis of efficacy endpoints are addressed in general FDA guidance documents.[6] Issues that commonly arise in oncology applications are discussed in this guidance.

Table 1. A Comparison of Important Cancer Approval Endpoints

Endpoint	**Regulatory Nature of Evidence**	**Assessment**	**Some Advantages**	**Some Disadvantages**
Overall Survival	Clinical benefit	• Randomized studies needed • Blinding not essential	• Universally accepted direct measure of benefit • Easily measured • Precisely measured	• Requires larger studies • Requires longer studies • Potentially affected by crossover therapy • Does not capture symptom benefit • Includes noncancer deaths
Disease-Free Survival	Surrogate for accelerated approval or regular approval*	• Randomized studies needed • Blinding preferred	• Considered to be clinical benefit by some • Needs fewer patients and shorter studies than survival	• Not a validated survival surrogate in most settings • Not precisely measured; subject to assessment bias • Various definitions exist

*Adequacy as a surrogate endpoint for accelerated approval or regular approval is highly dependent upon other factors such as effect size, effect duration, and benefits of other available therapy. See text for details.

continued

---

[6] See ICH guidance for industry *E9 Statistical Principles for Clinical Trials* (http://www.fda.gov/cder/guidance/index.htm)

Table 1, continued

Endpoint	Regulatory Nature of Evidence	Assessment	Some Advantages	Some Disadvantages
Objective Response Rate (ORR)	Surrogate for accelerated approval or regular approval*	• Single-arm or randomized studies can be used • Blinding preferred in comparative studies	• Can be assessed in single-arm studies	• Not a direct measure of benefit • Usually reflects drug activity in a minority of patients • Data are moderately complex compared to survival
Complete Response (CR)	Surrogate for accelerated approval or regular approval*	• Single-arm or randomized studies can be used • Blinding preferred in comparative studies	• Durable CRs represent obvious benefit in some settings (see text) • Can be assessed in single-arm studies	• Few drugs produce high rates of CR • Data are moderately complex compared to survival
Progression Free Survival (PFS)	Surrogate for accelerated approval or regular approval*	• Randomized studies needed • Blinding preferred • Blinded review recommended	• Activity measured in responding and stable tumors • Usually assessed prior to change in therapy • Less missing data than for symptom endpoints • Assessed earlier and in smaller studies compared with survival	• Various definitions exist • Not a direct measure of benefit • Not a validated survival surrogate • Not precisely measured compared with survival • Is subject to assessment bias • Frequent radiologic studies are needed • Data are voluminous and complex compared to survival
Symptom Endpoints	Clinical benefit	• Usually needs randomized blinded studies (unless endpoints have an objective component and effects are large — see text)	• Direct measure of benefit	• Blinding is often difficult in oncology trials • Missing data are common • Few instruments are validated for measuring cancer-specific symptoms • Data are voluminous and complex compared to survival

*Adequacy as a surrogate endpoint for accelerated approval or regular approval is highly dependent upon other factors such as effect size, effect duration, and benefits of other available therapy. See text for details.
Abbreviations: complete response (CR); objective response rate (ORR); progression-free survival (PFS).

A. Overall Survival

Overall survival is defined as the time from randomization until death from any cause, and is measured in the intent to treat (ITT) population. Survival is the most reliable cancer endpoint, and when studies can be conducted to adequately assess it, it is usually the preferred endpoint. An improvement in survival is of unquestioned clinical benefit. The endpoint is precise and easy to measure, documented by the date of death. Bias is not a factor in endpoint measurement.

Overall survival almost always needs to be evaluated in randomized controlled studies. Historically controlled data are seldom reliable for time-dependent endpoints such as overall survival unless treatment effects are extreme (e.g., acute leukemia, testicular cancer). Apparent differences in outcome between historical controls and current treatment groups can arise from differences other than drug treatment, including patient selection, improved imaging techniques (which can alter tumor staging and prognosis), or improved supportive care. Randomized studies minimize the effect of such differences by allowing a comparison of outcomes in patient groups where such factors should be similar. Demonstration of a statistically significant improvement in overall survival is usually considered to be clinically significant, and has often supported new drug approval.

Criticisms of survival as an endpoint stem not from doubts about the worth of a proven survival benefit, but from difficulties in performing studies large enough or long enough to detect a survival improvement, difficulties in determining a drug's effect on survival because of the confounding effects of subsequent cancer therapy, or a concern that the drug may be effective in only a small fraction of those treated, making it difficult to see an effect on survival in the whole population.

B. Endpoints Based on Tumor Assessments

In this section we discuss several endpoints that are based on tumor assessments and are therefore unique to oncology. These endpoints include disease-free survival, objective response rate, time to progression, progression-free survival, and time to treatment failure. The data collection and analysis of all time-dependent endpoints is complex, particularly when the assessments are indirect and based on calculations and estimates as is the case for tumor measurements. The discussion of progression-free survival data collection and analysis is particularly complex and is supplemented by tables in Appendix 3 of this guidance.

Selection of tumor-assessment endpoints for efficacy trials should include two judgments. First, will the endpoint support accelerated approval (is the endpoint a surrogate reasonably likely to predict clinical benefit and does the drug provide an advantage over available therapy) or regular approval (is it an established and/or validated surrogate for, or a direct measure of, clinical benefit)? Second, will the results be reliable, given the potential for uncertainty or bias in tumor endpoint assessments? Drug applications using studies that rely on tumor measurement based endpoints as sole evidence of efficacy should generally provide confirmatory evidence from a second trial. Both the precision and the clinical meaning of endpoints based on tumor assessments can vary in different cancer settings. For instance, response rate determinations in malignant mesothelioma and pancreatic cancer are often unreliable because of the difficulty in measuring these tumors with currently available imaging modalities.

When the primary study endpoint for drug approval is based on tumor measurements (e.g., progression-free survival or ORR), it is recommended that tumor endpoint assessments generally be verified by central reviewers blinded to study treatment (see Appendix 4), especially when the study itself cannot be blinded. Although the FDA will generally not ask that all tumor images be submitted with the marketing application, it may need to audit a sample of the scans to verify the central review process. In all cases, we recommend submitting primary electronic data documenting tumor measurements and assessments.[7] Additional details regarding data collection are listed in Appendix 1.

1. Disease-Free Survival

Disease-free survival (DFS) is usually defined as the time from randomization until recurrence of tumor or death from any cause. Although DFS can also be an important endpoint when a large percentage of patients achieve complete responses with chemotherapy, the most frequent use of this endpoint is in the adjuvant setting after definitive surgery or radiotherapy. In either of these settings, DFS has special meaning to patients because until a recurrence occurs, a patient can hope for cure. Whereas overall survival is the standard endpoint for most adjuvant settings, DFS has been the primary basis of approval for hormonal therapy after initial surgery for breast cancer. An important consideration is whether prolongation of DFS represents intrinsic benefit or only a potential surrogate for survival prolongation. In December 2003, the consensus of the ODAC was that prolongation of DFS represented clinical benefit, but that the magnitude of this benefit should be carefully weighed against the toxicity of adjuvant treatment, particularly as measured by effects on patient function. In May 2004, the ODAC recommended that DFS be considered an acceptable endpoint for colon cancer drugs in the surgical adjuvant setting, provided certain conditions were met.[8] Additional cancer-specific guidances will address the acceptability of DFS in other cancer settings.

Important considerations in evaluating DFS as a potential endpoint include the estimated size of the treatment effect, proven benefits of standard therapies, and details of trial design. For instance, when a new drug is compared to a control drug that is known to improve overall survival, an important consideration is whether the DFS of the new drug is superior to, or only noninferior to, the control. Clearly, proof of superiority with regard to a surrogate endpoint is more persuasive than a demonstration of noninferiority. Furthermore, relying on a conclusion of noninferiority based on a surrogate endpoint to support a conclusion of noninferiority with respect to the definitive endpoint is problematic. Another critical issue is whether the duration of study follow-up is adequate to evaluate the durability of the DFS benefit.

We suggest that the protocol carefully detail both the definition of DFS and the schedule for follow-up studies and visits. Unscheduled assessments can occur for many reasons (including tumor-related symptoms, drug toxicity, anxiety), and differences between study arms in the frequency or reason for unscheduled assessments is likely to introduce bias. This potential bias can be minimized by blinding patients and investigators to the treatment assignments if feasible. The potential effects of bias due to unscheduled assessments can be evaluated by comparing their frequency between treatment arms and by performing statistical analyses that assign events from unscheduled visits to the time of the next scheduled visit.

---

[7] See guidance for industry *Cancer Drug and Biological Products — Clinical Data in Marketing Applications* (http://www.fda.gov/cder/guidance/index.htm)

[8] Transcripts are available at http://www.fda.gov/cder/drug/cancer_endpoints/default.htm.

Another issue in defining DFS is whether deaths occurring without prior documentation of tumor progression should be scored as DFS events (disease recurrences) or should be censored in the statistical analysis. All methods for statistical analysis of deaths have limitations. The approach that seems less prone to introducing bias is to consider all deaths as recurrences. Limitations of this approach are a potential decrease in statistical power of the study (by *diluting* the cancer-related events with deaths not related to cancer) and a potential to falsely prolong the DFS estimates in patients who die after a long unobserved period. The latter could introduce bias if the frequency of long-term follow-up visits is dissimilar on the study arms or if there is nonrandom dropout due to toxicity. Some analyses count cancer-related deaths as DFS events and censor noncancer deaths. This method has the potential for bias in the post hoc determination of the cause of death. Furthermore, any method that censors patients, whether at death or at the last visit, assumes that the censored patients have the same risk of recurrence as noncensored patients. This critical assumption needs close examination in any setting where deaths are to be censored. In settings where deaths due to causes other than cancer are common (e.g., studies of patients with early metastatic prostate cancer), censoring deaths can be appropriate.

2. Objective Response Rate

ORR is the proportion of patients with tumor shrinkage of a predefined amount lasting for a predefined minimum period of time. Response duration is usually measured from the time of initial response until documented tumor progression. The FDA has generally defined ORR as the sum of partial responses plus complete responses. When defined in this manner, ORR is a measure of drug antitumor activity even in a single-arm study. Some sponsors have proposed including stable disease as a component of ORR; however, evaluating drug effects based on the stable disease rate generally involves comparison to a randomized concurrent control. Also, stable disease incorporates components of time to progression or progression-free survival, which can be captured in a separate measurement. A variety of response criteria have been considered appropriate, including the RECIST criteria (Therasse and Arbuck et al., 2000, New Guidelines to Evaluate Response to Treatment in Solid Tumors, J Natl Cancer Inst, 92:205-16). Important issues for determining the clinical and regulatory significance of ORR include response duration, the percentage of complete responses, the toxicity of treatment, and associated improvement in tumor-related symptoms. These issues, in addition to an assessment of benefits of existing therapies, determine whether ORR will support marketing authorization, either for regular approval (as a full surrogate for clinical benefit) or for accelerated approval (as a *reasonably likely surrogate*).

It is important that criteria for response and progression be detailed in the protocol, and data should be carefully and completely collected at intervals specified in the protocol.

3. Time to Progression and Progression-Free Survival

In the past, time to progression (TTP) (the time from randomization until objective tumor progression) and progression-free survival (PFS) (the time from randomization until objective tumor progression or death) have seldom served as primary endpoints for drug approval. Time to symptomatic progression, which would represent a clear clinical benefit, is infrequently assessed but would be a credible endpoint of a well-conducted (generally blinded) trial. In December 2003, the ODAC discussed both potential roles of TTP and PFS in cancer drug approval and the committee's preference for PFS versus TTP.[9] The ODAC suggested relying on these endpoints in selected clinical situations, such as diseases with low complete response rates or when documentation of a survival benefit in clinical trials can be difficult. In settings where most patients are symptomatic, the ODAC preferred measuring tumor response and symptom benefit. The definition of tumor progression varies widely; therefore, it is important that it be carefully detailed in the protocol.

a. TTP vs. PFS

The ODAC consensus was that PFS is a better predictor of clinical benefit than TTP and thus preferable as a drug approval endpoint when used as a surrogate for clinical benefit (rather than just as an indicator of antitumor activity) because PFS includes deaths. Unanticipated effects of drugs on survival would thus be included in the endpoint. In the analysis of TTP, deaths are censored, either at the time of death or at an earlier visit. This approach is questionable because it can represent *informative censoring* (i.e., there may be a nonrandom pattern of loss from the study). It seems unlikely in most cancer settings that patient deaths are randomly related to tumor progression (e.g., it is likely that some deaths result from complications of undocumented cancer progression). Therefore, in most settings PFS is the preferred regulatory endpoint. In settings where most deaths are due to causes other than cancer, however, TTP can be an appropriate endpoint.

b. PFS as an endpoint to support drug approval

Some advantages and disadvantages of using PFS as an endpoint to support cancer drug approval are listed in Table 1. Conceptually, PFS has desirable qualities of a surrogate endpoint because it reflects tumor growth (a phenomenon likely to be on the causal pathway for cancer-associated morbidity and death), can be assessed prior to demonstration of a survival benefit, and is not subject to the potential confounding impact of subsequent therapy (unless worsening of a blood marker leads to a change in treatment prior to progression). Moreover, an effect on PFS occurs earlier than an effect on survival, so that a given advantage, say a median improvement of 3 months, represents a larger (and thus more detectable) hazard ratio improvement than would a 3-month median survival benefit occurring later. The formal validation of PFS as a surrogate for survival for the many different malignancies that exist, however, would be difficult. Data are usually insufficient to allow a robust evaluation of the correlation between effects on survival and PFS. Oncology trials are often small, and proven survival benefits of existing drugs are generally modest. The role of PFS as an endpoint to support licensing approval varies in different cancer settings. In some settings PFS prolongation might be an accepted surrogate endpoint for clinical benefit to support regular approval, and in others it may be a surrogate reasonably likely to predict benefit for accelerated approval. Important considerations will be the magnitude of the effect, the toxicity profile of the treatment, and the clinical benefits and toxicities of available therapies. These issues will be discussed in future guidance documents for specific cancer settings.

---

[9] Transcripts are available at http://www.fda.gov/cder/drug/cancer_endpoints/default.htm.

c. PFS trial design issues

It is important that methodology for assessing, measuring, and analyzing PFS be detailed in the protocol and statistical analysis plan. It is also important to carefully define tumor progression criteria in the protocol. There are no standard regulatory criteria for defining progression. Sponsors have used a variety of different criteria, including the RECIST criteria. The broad outline presented in most published PFS criteria should be supplemented with additional details in the protocol and statistical analysis plan. It is important that visits and radiological assessments be symmetric on the two study arms to prevent systematic bias. When possible, studies should be blinded. Blinding is particularly important when patient or investigator assessments are included as components of the progression endpoint. It is important that the FDA and the sponsor agree prospectively on the protocol, data to be recorded on the case report form, statistical analysis plan (including analysis of missing data and censoring methods), and, if applicable, the operating procedures of an independent endpoint review committee (discussed in Appendix 4). The effect of follow-up visit frequency has been debated. Frequent regular assessments, depending on the type and stage of cancer, ensure that most progression events will be detected on radiologic scans rather than as symptomatic events. This approach increases the expense and difficulty of the study, including an increased data collection burden on the investigator and an increased number of scans for patients, and may not mirror clinical practice standards.

d. Analysis of PFS

The analysis of PFS is complicated by missing data. It is important that the protocol specify what constitutes an adequate assessment visit for each patient (i.e., a visit when all scheduled tumor assessments have been done). The analysis plan should outline a comparison of the adequacy of follow-up in each treatment arm and specify how incomplete or missing follow-up visits will be handled with regard to censoring. For instance, if one or more assessment visits are missed just prior to the progression event, to what date should the progression event be assigned? It is important that the analysis plan specify the primary analysis and one or more sensitivity analyses. For instance, in the previous example, the primary analysis might assign the actual date of observed progression as the progression date. The sensitivity analysis might censor the data at the last adequate assessment visit. Although both analyses are problematic (the best solution to missing data is to have none), the conclusion is probably valid if it is supported by the results of both the primary and the sensitivity analyses. Other methods could be considered if adequately supported by the sponsor. The analysis plan should evaluate the number of deaths in patients who have been lost to follow-up for more than a substantial (prespecified) time. An imbalance in such deaths could bias the measurement of PFS, artificially prolonging PFS on the arm with less adequate follow-up.

Because progression data can be collected from a variety of sources (including physical exams at unscheduled visits and radiologic scans of various types) and at a variety of times, it is important that data collection efforts for each assessment visit be limited to a specified short time interval prior to the visit. When data are collected over a longer time, the question then arises: What date should serve as the progression date or the censoring date? A common method is to assign progression to the earliest observed time when an observation shows progression and to censor at the date when the last radiologic assessment determined a lack of progression. Because this method could introduce an assessment bias, especially in unblinded trials, we recommend assigning the progression and censoring times to the time of the scheduled assessment visits. A study of time to symptomatic progression, if conducted blindly and with few scheduled assessments, in contrast, could use the actual time of observed symptom progression. The PFS date based on a death, however, would be the date of death rather than the assigned visit date since death ascertainment is not related to visit time and not subject to interpretation.

Appendix 3 provides a set of tables for potential analyses of PFS that could be used for primary or sensitivity analyses. We recommend that plans for PFS data collection and analysis be discussed with the FDA at end-of-phase 2 meetings and verified in special protocol assessments.

e. Future methods for assessing progression

In the future, it is important that other methods of progression assessment be evaluated as potential surrogate endpoints for regular approval or accelerated approval. One proposed method (not used to date) is the single time point assessment which could decrease the complexity of progression assessment and eliminate time-dependent assessment bias. In the single time point analysis, progression would be assessed at baseline and at one prespecified time after randomization. If patients progress prior to the specified time, radiologic scans could document progression and the patient could go off-study. All other patients would have a detailed radiologic evaluation at the prespecified follow-up time. The statistical analysis could compare the proportions of patients on each study arm with progression on or before the prespecified time after randomization. Potential problems with this approach are decreased statistical power, potential for missing a small benefit at a time different from the prespecified time, and lack of information regarding the relationship between the single time point analysis and the familiar endpoints of progression-free survival and overall survival. Although this approach could provide some advantages and decrease assessment bias, study dropouts prior to progression could present the same difficulty as they do for all progression endpoints. Settings in which further evaluation of this approach seems warranted are those where a significant and durable effect on progression-free survival is expected and where complete progression-free survival data collection seems impossible or impractical.

4. Time to Treatment Failure

Time to treatment failure (TTF) is a composite endpoint measuring time from randomization to discontinuation of treatment for any reason (including progression of disease, treatment toxicity, and death). Defined that way, TTF is not recommended as an endpoint for drug approval because it combines efficacy and toxicity measures. For example, suppose the standard comparator (Drug A) provides a known survival benefit, but only at the cost of considerable toxicity with many patients leaving therapy because of that toxicity. A nontoxic investigational drug (Drug B) could have a significantly longer TTF than Drug A solely because it caused fewer toxic dropouts. These data alone could not support drug approval because they would not demonstrate that Drug B is effective. Drug approval would require a demonstration of Drug B efficacy, such as a survival improvement or other clinical benefit.

C. Endpoints Involving Symptom Assessment

Symptomatic improvement has always been considered a clinical benefit, and many FDA cancer drug approvals have used patient symptom assessments and/or physical signs thought to represent symptomatic improvement (e.g., weight gain, decreased effusion) as the primary evidence of effectiveness. To date, broader measures of health-related quality of life (HRQL instruments) have not served this role. HRQL is discussed in a separate FDA draft guidance on patient-reported outcomes (PRO).[10] The FDA has relied on symptom scores, signs, and symptoms representing obvious benefit (e.g., decreased esophageal obstruction, fewer bone fractures, reduced size and number of skin lesions, physician actions [need for radiation therapy in response to painful bone metastases], physician assessments of performance status, and patient-reported assessments of symptom scales). Relying on such evidence of clinical benefit as the basis for approval has allowed the FDA to approve cancer drugs earlier than if demonstration of a survival benefit had been required. It seems self-evident that cancer patients will be in most cases the best source for determining effects on patient symptoms, so that PRO instruments seem most appropriate. Formal PRO instruments can be designed that focus on specific symptoms (e.g., a pain scale) or on a broader array of physical, emotional, and activity measures.

The use of improvement of signs and symptoms or QOL assessments as primary endpoints to support cancer drug approval requires discrimination between tumor symptoms and drug toxicity, especially when evidence is based on comparison to a toxic active control. This poses particular problems for general HRQL scales, which, by definition, are multidimensional scales including elements other than physical problems. An apparent effectiveness advantage of one drug over another measured on a global HRQL instrument might simply indicate less toxicity of one product or regimen versus the other, a matter of interest but not an effectiveness measure. Morbidity endpoints used to date for cancer drug approvals have possessed *face validity* (value obvious to patients and physicians, for example, an endpoint based on functional measures such as the ability to swallow solids, liquids, or nothing) and have not measured benefit and toxicity on the same scale.

1. Specific Symptom Endpoints

One endpoint the FDA has suggested to sponsors is *time to progression of cancer symptoms,* an endpoint similar to time to progression. This endpoint would be a direct measure of clinical benefit rather than a potential surrogate. Sponsors have cited several problems with this approach. First, because few cancer trials are blinded, assessments can be biased and therefore unreliable. Another problem is the usual delay between tumor progression and the onset of cancer symptoms. Often alternative treatments are begun before reaching the symptom endpoint, which can confound the results. Many cancer trials are performed in patients with little prior exposure to chemotherapy and who usually have minimal cancer symptoms. Finally, it can sometimes be difficult to differentiate tumor symptoms from drug toxicity, a problem noted in discussions of time to treatment failure and HRQL. *Time to progression of symptoms* and *time to onset of symptoms* can be reasonable endpoints in cancer settings where treatment can be blinded, most progressing patients are symptomatic, no effective therapy exists, and less frequent radiologic follow-up is appropriate. Symptom data should be carefully collected using a validated instrument according to a schedule detailed in the protocol.

---

[10] The draft guidance for industry *Patient-Reported Outcome Measures: Use in Medical Product Development to Support Claims* is currently being developed and is expected to publish in the summer of 2005. When final, this guidance will represent the FDA's current thinking on this topic. For the most recent version of a CDER or CBER guidance, check the CDER guidance Web page at http://www.fda.gov/cder/guidance/index.htm and the CBER Web page at http://www.fda.gov/cber/guidance/index.htm.

A *composite symptom endpoint* can be appropriate when the benefit of a drug is multifaceted. It is important that the components of the endpoint be related and generally of similar clinical importance. Drugs have been approved for treatment of patients with cancer metastases to the skeleton based on a composite benefit endpoint consisting of one or more skeletal-related event (SRE) that would be anticipated to be associated with pain and other distress. SREs are defined as pathologic fractures, radiation therapy to bone, surgery to bone, and spinal cord compression. Clinical Benefit Response, a composite endpoint of pain and analgesic consumption reported by the patient, and performance status assessed by a physician, in part supported approval of a drug to treat pancreatic cancer.

Selection of the appropriate population for study can be critical for documenting symptom benefit. Patients symptomatic at study baseline can be evaluated with a categorical symptom response analysis. This approach can be appropriate for diseases such as lung cancer, when most patients have symptoms at diagnosis. Studies of asymptomatic patients could use a time-to-first-symptom analysis. Even if the patient discontinues the study drug or begins a new drug, symptomatic progression could still be assessed if follow-up is continued until documentation of the first symptom. This approach is worth considering but has been infrequently attempted.

2. Problems Encountered with Symptom Data

Many problems have been encountered in the analysis of symptom data submitted to the FDA. The most important problem in oncology is that few trials are blinded so that the possibility of observer bias is difficult to exclude. Missing data are common and often cast doubt on study conclusions. It is critically important to have frequent assessments to minimize long unobserved gaps. In addition, symptom severity should be addressed, rather than providing only a binary present or absent. Withdrawing treatment because of drug toxicity or tumor progression is one cause of missing symptom data. Ideally, when patients stop treatment, data collection forms should continue to gather information to inform the analysis. Symptom data could lead to a large number of different endpoints, and prospectively defined statistical plans need to correct for multiplicity if each symptom is treated as a separate endpoint.

D. Biomarkers

To date, evidence from biomarkers assayed from blood or body fluids has not served as primary endpoints for cancer drug approval, although paraprotein levels measured in blood and urine have contributed to response endpoints for myeloma. Further research is needed to establish the validity of the available tests and determine whether improvements in such biomarkers are reasonably likely to predict clinical benefit (accelerated approval) or are established surrogates for clinical benefit (regular approval).

Although tumor markers are not yet used alone as a basis for marketing approval, the FDA has sometimes accepted their inclusion as elements in composite endpoints. For instance, women with ovarian cancer often show clinical deterioration from progression of unmeasured tumor. In blinded randomized controlled trials in advanced refractory ovarian cancer, the FDA has accepted use of a composite endpoint that included CA-125. The occurrence of certain clinical events (a significant decrease in performance status, or bowel obstruction) coupled with marked increases in CA-125 was considered progression in these patients. The use of prostate specific antigen (PSA) was discussed at a recent workshop on prostate cancer endpoints. Different methods of evaluating PSA as an endpoint were discussed, including PSA response, PSA slope, and PSA velocity. Although the FDA has not yet accepted a PSA endpoint to support drug approval, evaluation of additional data and further discussions of PSA endpoints are planned in future workshops and ODAC meetings.[11]

IV. ENDPOINTS AND CLINICAL TRIAL DESIGN; SELECTED ISSUES

By law, the FDA must base new drug approval decisions on substantial evidence of efficacy from “adequate and well-controlled investigations.” Regulations describe the meaning of “adequate and well-controlled investigations.” Studies must allow a valid comparison to a control and must provide a quantitative assessment of the drug’s effect. (See 21 CFR 314.126.) Below we discuss several issues related to the design of cancer trials intended to support drug approval.

A. Single-Arm Studies

The most reliable method for demonstrating efficacy is to show a statistically significant improvement in a clinically meaningful endpoint in blinded randomized controlled trials. Other approaches have also been successful in certain settings. In settings where there is no effective therapy and where major tumor regressions can be presumed to occur infrequently in the absence of treatment (a historical control), the FDA has sometimes accepted ORR and response duration observed in single-arm studies as substantial evidence supporting accelerated approval or even regular approval (e.g., when many complete responses were observed or when toxicity was minimal or modest). In contrast to the success of this approach, evidence from historically controlled trials attempting to show improvement in time-to-event endpoints such as survival, time to progression, or progression-free survival have seldom been persuasive support for drug approval, except when treatment provides survival outcomes that contrast markedly with historical experience (e.g., testicular cancer, acute leukemias). In most cases, however, these outcomes vary among study populations in ways that cannot always be predicted; for example, changes in concomitant supportive care or frequency and method of tumor assessment can differ by location or change over time. Consequently, comparisons involving these time-to-event endpoints generally need a concurrent control (preferably in a randomized trial), unless, as noted, the effect is very large.

---

[11] Transcripts are available at http://www.fda.gov/cder/drug/cancer_endpoints/default.htm.

B. Studies Designed to Demonstrate Noninferiority

The goal of noninferiority (NI) trials is to demonstrate the effectiveness of a new drug showing that it is not less effective, by a predefined amount, than a standard regimen known to have the effect being investigated (Temple and Ellenberg, 2000, Placebo-Controlled Trials and Active-Control Trials in the Evaluation of New Treatments, Part 1: Ethical and Scientific Issues, Ann Intern Med, 2000 Sep 19; 133(6):455-63).[12] The difference to be ruled out, the *noninferiority margin*, cannot be larger than the effect of the control drug in the new study. As that effect is not measured (the new study does not have a no-treatment arm), the effect must be assumed based on the previous studies of the control drug that documented its effect. If the new drug is inferior by more than the noninferiority margin, it would have no effect at all. In most cases the NI margin is not set at the control drug's full effect, but at some fraction of it (e.g., 50 percent), so that the study seeks to show that at least 50 percent of the control drug effect is preserved.

There are multiple difficulties with NI trials. NI trials rely on historical data to establish the expected size of treatment effect of the active control. In many situations adequate historical data for the control do not exist. Moreover, a critical assumption is that the treatment effect of the active control that was observed historically will also be observed in the current population in the new study. This assumption is difficult to support, as results of trials are almost never identical (although one can evaluate control regimen response rates in the historical and NI trial populations as some measure of comparability). Optimally, the estimated size of the treatment effect of the active control would be based on a comprehensive meta-analysis of historical studies that reproducibly demonstrate the effectiveness, compared to no treatment, of the control agent. In the oncology setting, however, information is often lacking on effects compared to a no-treatment control. The variability in the meta-analysis will be reflected in the choice of the noninferiority margin. But there may be little data from randomized controlled trials available to estimate the treatment effect and thus no basis for estimating the control treatment effect. Furthermore, subsequent events in the trial, especially crossover from the control, can invalidate NI survival analyses (producing a bias toward a showing of no difference). NI designs generally require many patients in order to provide meaningful results. Given the complex issues involved, we strongly recommend that sponsors designing noninferiority trials consult early with the FDA. Because of the difficulties with the design, conduct, and analysis of NI trials, a single NI trial seldom provides sufficient evidence of efficacy to support drug approval.

When the new treatment has a different toxicity profile from available treatments, it may be possible to *design around* the NI study problem by conducting an *add-on* study, adding new drug or placebo/no treatment to the standard therapy. This will not be possible if the goal is to show a new treatment to be less toxic than existing therapy (but still effective). In this case the NI design is unavoidable in order to demonstrate that the survival benefit of the standard drug is retained by the experimental drug. If the standard drug is associated with only a small proven survival benefit, however, interpretation of an NI study is difficult or impossible. Moreover, the size of such NI trials can be prohibitively large.

---

[12] See ICH guidance for industry *E10 Choice of Control Group and Related Issues in Clinical Trials* (http://www.fda.gov/cder/guidance/index.htm)

C. No Treatment or Placebo Control

Giving no anticancer drug treatment to patients in the control arm of a cancer study is often considered unethical, but, in some settings, it can be acceptable. For instance, in early stage cancer when standard practice is to give no treatment, comparison of a new agent to a no-treatment control would be acceptable. This approach would not be an ethical problem in the so-called *add-on* design, when all patients receive standard treatment plus either no additional treatment or the experimental drug. Using a control group that receives only best supportive care is acceptable in an advanced refractory setting where there is no effective therapy. Placebos (identical appearing inactive controls) are generally preferred to no-treatment controls because they permit blinding. With many cytotoxic cancer drugs, blinding may not be feasible because of a relatively high rate of recognizable toxicities, but newer interventions, many of them much less toxic, are increasingly being studied in blinded trials.

D. Isolating Drug Effect in Combinations

Because marketing approval is usually for a single drug product rather than for a drug combination, clinical trials supporting regulatory approval need to isolate the effectiveness of the proposed agent. Evidence is needed showing not only the effectiveness of the regimen but also establishing the contribution of the new drug to that regimen. One way to demonstrate the individual contribution of a new drug in a regimen is using the *add-on* design previously discussed. Sometimes the clinical effects seen in early phases of development can be used to establish the contribution of a drug to a drug regimen, particularly if the combination is more effective than any of the individual components. We recommend discussing these issues with the FDA at end-of-phase 1 or end-of-phase 2 meetings.

E. Trial Designs for Radiotherapy Protectants and Chemotherapy Protectants

Radiotherapy protectants and chemotherapy protectants are drugs designed to ameliorate the toxicities of radiotherapy or chemotherapy. Trials to evaluate these agents usually have two objectives. The first is to assess whether the protecting drug achieves its intended purpose of ameliorating the cancer treatment toxicity. Unless the mechanism of protection is clearly unrelated to the mechanism of antitumor activity (e.g., antiemetic agents which ameliorate nausea via central nervous system receptors), a second trial objective is to determine whether anticancer efficacy is compromised by the protectant. Because the comparison of antitumor activity between the two arms of the trial is a noninferiority comparison, a large number of patients may be required to achieve this objective. Generally, a second study is needed to confirm the findings. A critical question for the future is whether, in such cases where the same drug is studied in both arms, ORR should be considered a sufficient endpoint for comparing drug activity and benefit.

V. SUMMARY AND CONCLUSION

Although general principles outlined in this guidance should help sponsors select endpoints for marketing applications, we recommend that sponsors meet with the FDA before submitting protocols intended to support NDA or BLA marketing applications. The FDA will ensure that these meetings include a multidisciplinary FDA team of oncologists, statisticians, clinical pharmacologists, and often external expert consultants. Sponsors may submit protocols after these meetings and request a special protocol assessment that provides the acceptability of endpoints and protocol design to support drug marketing applications.[13] Ultimately, of course, marketing approval will depend not only on the design of a single trial, but on FDA review of the results and data from all studies in the drug marketing application.

---

[13] See guidance for industry *Special Protocol Assessment* (http://www.fda.gov/cder/guidance/index.htm)

APPENDIX 1:
THE COLLECTION OF TUMOR MEASUREMENT DATA[14]

The following are important considerations for tumor measurement data. The Agency recommends that:

- The case report form (CRF) and electronic data document the target lesions identified during the baseline visit prior to treatment. Retrospective identification of such lesions would rarely be considered reliable.
- Tumor lesions are assigned a unique identifying letter or number. This allows differentiating among multiple tumors occurring at one anatomic site and matching of tumors measured at baseline and tumors measured during follow-up.
- A mechanism ensures complete collection of data at critical times during follow-up. It is important that the CRF ensures that all target lesions are assessed at each follow-up visit and that all required follow-up tests are done with the same imaging/measuring method.
- The CRF contains data fields that indicate whether scans were performed at each visit.
- A zero is recorded when a lesion has completely resolved. Otherwise, disappearance of a lesion cannot be differentiated from a missing value.
- Follow-up tests allow timely detection of new lesions both at initial and new sites of disease. It is important that the occurrence of and location of new lesions be recorded in the CRF and the submitted electronic data.

---

[14] *Tumor data* in this section refers to data in SAS transport files, not images. Images are not generally submitted to the NDA/BLA, but may be audited by the FDA during the review process.

APPENDIX 2:
ISSUES TO CONSIDER IN PFS ANALYSIS

The protocol and statistical analysis plan (SAP) of a study should detail the primary analysis of progression-free survival (PFS). This includes a detailed description of the endpoint, acceptable modalities for evaluating tumors, and procedures for minimizing bias when determining progression status, such as procedures for an independent endpoints review committee. It is important that one or two secondary analyses be specified to evaluate anticipated problems in trial conduct and to assess whether results are robust. The following are several important factors to consider.

- **Definition of progression date.** Survival analyses use the exact date of death. In analyses of PFS, however, the exact progression date is unknown. The following are two methods for defining the *recorded progression date (PDate)* used for PFS analysis.

  1. One approach assigns PDate to the first time at which progression can be declared:
     - For progression based on a new lesion, the PDate is the date of the first observation that detects the new lesion.
     - For progression based on the sum of target lesion measurements, PDate is the date of the last observation or radiologic assessment of target lesions (if multiple assessments are done at different times).

     This approach can introduce between-arm bias if radiologic assessments are done earlier or more frequently in one treatment arm.

  2. A second approach assigns the PDate to the date of the scheduled clinic visit immediately after all radiologic assessments (which collectively document progression) have been done. Although this approach provides a less accurate estimate of the true date of progression, the error should be symmetrically distributed between arms, and between-arm bias is minimized.

- **Definition of censoring date.** Censoring dates are defined in patients with no documented progression prior to data cutoff or dropout. In these patients, the censoring date is often defined as the last date on which progression status was adequately assessed. One acceptable approach uses the date of the last assessment performed. However, multiple radiologic tests can be evaluated in the determination of progression. A second acceptable approach uses the date of the clinic visit corresponding to these radiologic assessments.

- **Definition of an adequate PFS evaluation.** In patients with no evidence of progression, censoring for PFS often relies on the date of the last *adequate tumor assessment.* A careful definition of what constitutes an adequate tumor assessment includes adequacy of target lesion assessments and adequacy of radiologic tests both to evaluate nontarget lesions and to search for new lesions.

- **Analysis of partially missing tumor data.** Analysis plans should describe the method for calculating progression status when data are partially missing from *adequate tumor* *assessment* visits. For instance, are the values for missing target lesions to be *carried forward*?

- **Completely missing tumor data.** Assessment visits where no data are collected are sometimes followed by death or by assessment visits showing progression; in other cases the subsequent assessment shows no progression. In the latter case, at first glance, it might seem acceptable to continue the patient on study and continue monitoring for evidence of progression. This approach, however, treats missing data differently depending upon subsequent events and could represent informative censoring. Therefore, another possibility is for the primary analysis to include data from subsequent PFS assessments when only a single follow-up visit is missed but censor data when there are two or more missed visits. It is important that the SAP detail primary and secondary PFS analyses to evaluate the potential effect of missing data. Reasons for dropouts should be incorporated into procedures for determining censoring and progression status. For instance, for the primary analysis, patients going off-study for undocumented clinical progression, change of cancer treatment, or decreasing performance status could be censored at the last adequate tumor assessment. The secondary sensitivity analysis would include these dropouts as progression events.

- **Progression of nonmeasurable disease.** When appropriate, progression criteria should be described for each assessment modality (e.g., CT scan, bone scan). It is important that scans documenting progression based on nonmeasurable disease be verified by a blinded review committee and be available for verification by the FDA if needed.

- **Suspicious lesions.** Sometimes new lesions are identified as suspicious. An algorithm should be provided for following up these lesions and for assignment of progression status at the time of analysis. For example, a radiological finding identified as suspicious at visit one might be verified as being a new tumor at visit three. It is important that the protocol or analytical plan clarify whether the progression time would be visit one or visit three.

APPENDIX 3:
EXAMPLE TABLES FOR PFS ANALYSIS

As discussed in Section III.B., sensitivity analyses may be helpful in determining whether the PFS analysis is robust. Different sensitivity analyses can be described in tables that specify how to assign dates of progression events and dates for censoring of progression data. The following three tables describe examples of three different sensitivity analyses:

a. Table A represents a sensitivity analysis that only includes well-documented and verifiable progression events. Other data are censored. In Table A the progression dates are:

- Based only on radiologic assessments verified by an independent review committee (IRC). *Clinical progression* is not considered a progression endpoint.
- Assigned to the first time when tumor progression was noted.
- The date of death when the patient is closely followed. Deaths occurring after two or more missed visits, however, are censored at last visit.

Table A. PFS 1 (includes documented progression only)

Situation	**Date of Progression or Censoring**	**Outcome**
No baseline tumor assessments	Randomization	Censored
Progression documented between scheduled visits	Earliest of: • Date of radiologic assessment showing new lesion (if progression is based on new lesion); or • Date of last radiologic assessment of measured lesions (if progression is based on increase in sum of measured lesions)	Progressed
No progression	Date of last radiologic assessment of measured lesions	Censored
Treatment discontinuation for undocumented progression	Date of last scan of measured lesions	Censored
Treatment discontinuation for toxicity or other reason	Date of last radiologic assessment of measured lesions	Censored
New anticancer treatment started	Date of last radiologic assessment of measured lesions	Censored
Death before first PD assessment	Date of death	Progressed
Death between adequate assessment visits	Date of death	Progressed
Death or progression after more than one missed visit	Date of last radiologic assessment of measured lesions	Censored

The sensitivity analysis in Table B corrects for potential bias in follow-up schedules for tumor assessment by assigning the dates for censoring and events only at scheduled visit dates.

Table B. PFS 2 (uniform progression and assessment dates)

Situation	Date of Progression or Censoring	Outcome
No baseline tumor assessments	Randomization	Censored
Progression documented between scheduled visits	Date of next scheduled visit	Progressed
No progression	Date of last visit with adequate assessment	Censored
Treatment discontinuation for undocumented progression	Date of last visit with adequate assessment	Censored
Treatment discontinuation for toxicity or other reason	Date of last visit with adequate assessment	Censored
New anticancer treatment started	Date of last visit with adequate assessment	Censored
Death before first PD assessment	Date of death	Progressed
Death between adequate assessment visits	Date of death	Progressed
Death or progression after more than one missed visit	Date of last visit with adequate assessment	Censored

b. The sensitivity analysis in Table C evaluates PFS according to the investigator's assessment.

Table C. PFS 3 (includes investigator claims)

Situation	Date of Progression or Censoring	Outcome
No baseline assessment	Randomization	Censored
Progression documented between scheduled visits	Next scheduled visit	Progressed
No progression	Date of last visit with adequate assessment	Censored
Investigator claim of clinical progression	Scheduled visit (or next scheduled visit if between visits)	Progressed
Treatment discontinuation for toxicity or other reason	Date of last visit with adequate assessment	Censored
New anticancer treatment started with no claim of progression	Date of last visit with adequate assessment	Censored
Death before first PD assessment	Date of death	Progressed
Death between adequate assessment visits or after patient misses one assessment visit	Date of death	Progressed
Death after an extended lost-to-follow-up time (two or more missed assessments)	Last visit with adequate assessment	Censored

APPENDIX 4:
INDEPENDENT REVIEW OF TUMOR ENDPOINTS

Sponsors and the FDA need to be able to verify clinical trial results that support drug approval, including ORR and progression-free survival. ORR determined in single-arm studies can be verified by scrutiny of a limited number of images. However, when drug approval is based on measurement of progression-free survival in a randomized study, careful planning is needed to minimize bias and to allow the sponsor and the FDA to verify results. This is especially true when investigators and patients cannot be blinded to treatment assignment because of drug toxicities or manner of administration. An independent endpoints review committee (IRC) provides a mechanism to minimize bias in interpretation of the radiologic findings and independent adjudication of endpoints. We recommend that a clearly described written plan outlining the IRC function and process, sometimes called an independent review charter, be agreed upon with the FDA prior to study conduct. It is important that the plan describe how the independence of the committee will be assured; how images will be collected, stored, transported, and reviewed; how differences in image interpretation will be resolved; how clinical data will be used in final endpoint interpretation; and how, if needed, images and IRC results will be made available to the FDA for audit. The use of an IRC is discussed further in a draft guidance for the development of medical imaging products.[15]

---

[15] See draft guidance for industry *Developing Medical Imaging Drug and Biological Products, Part 3: Design, Analysis, and Interpretation of Clinical Studies*. When final, this guidance will represent the FDA's current thinking on this topic. For the most recent version of a CBER guidance, check the CBER guidance Web page at http://www.fda.gov/cber/guidelines.htm.